Toward a Global Maximization of the Molecular Similarity Function: Superposition of Two Molecules

PERE CONSTANS, LLUÍS AMAT, RAMON CARBÓ–DORCA

Institute of Computational Chemistry, University of Girona, Albereda 3-5, 17071 Girona, Catalonia, Spain

Received 7 May 1996; accepted 5 July 1996

ABSTRACT: A quantum similarity measure between two molecules is normally identified with the maximum value of the overlap of the corresponding molecular electron densities. The electron density overlap is a function of the mutual positioning of the compared molecules, requiring the measurement of similarity, a solution of a multiple-maxima problem. Collapsing the molecular electron densities into the nuclei provides the essential information toward a global maximization of the overlap similarity function, the maximization of which, in this limit case, appears to be related to the so-called *assignment problem*. Three levels of approach are then proposed for a global search scanning of the similarity function. In addition, atom–atom *similarity Lorentzian potential* functions are defined for a rapid completion of the function scanning. Performance is tested among these three levels of simplification and the Monte Carlo and simplex methods. Results reveal the present algorithms as accurate, rapid, and unbiased techniques for density-based molecular alignments.
 J Comput Chem **18:** 826–846, 1997

Keywords: molecular quantum similarity measures (MQSM); atomic shell approximation (ASA); global maximization; molecular alignments

Correspondence to: P. Constans
Contract grant sponsor: Catalan Government: contract grant number: CIRIT OA/au BQF93/24

 CCC 0192-8651 / 97 / 060826-21

Introduction

Molecular quantum similarity measures[1] (MQSM) based on first-order electron densities, $\rho(\mathbf{r})$, provides a rigorous basis for quantitative comparisons and superposition of molecules. Molecular superposition or alignments are applied in many structure problems, such as pharmacophore assessment,[2] quantitative comparison of stereochemistry,[3] crystal field distortion measures,[3] or pattern recognition in a 3-dimensional (3-D) structural data bases.[4] The usual computer programs available are based, for simplicity reasons, on an atom−atom distance minimization.[5] This technique implies assignments of equivalent atoms in the comparing molecules to be supplied as additional input data. A combinatorial problem appears when all possible equivalencies are considered, being treated with stochastic approximate algorithms.[6]

The MQSM are defined as the maximum of the similarity function,

$$z_{AB}(\mathbf{\Omega}) = \int\int \rho_A(\mathbf{r}_1)\mathbf{\Theta}(\mathbf{r}_1,\mathbf{r}_2)\rho_B(\mathbf{r}_2)\, d\mathbf{r}_1\, d\mathbf{r}_2. \quad (1)$$

In function (1), ρ_A and ρ_B are the electron densities of the pair of molecules A and B to be compared and $\mathbf{\Theta}(\mathbf{r}_1,\mathbf{r}_2)$ is a bielectron operator particularizing the similarity measure. The functional dependence of $z_{AB}(\mathbf{\Omega})$ is in respect to the set of translations and rotations specifying the mutual arrangement of the two molecules, indicated by $\mathbf{\Omega}$. A conceptually simple and structure-related similarity measure is found by taking $\mathbf{\Theta}(\mathbf{r}_1,\mathbf{r}_2)$ as the Dirac delta operator $\delta(\mathbf{r}_1 - \mathbf{r}_2)$; transforming integral (1) into an electron overlap function,

$$z_{AB}(\mathbf{\Omega}) = \int \rho_A(\mathbf{r})\rho_B(\mathbf{r})\, d\mathbf{r}. \quad (2)$$

The maximization of the overlap of the electron distributions leads to a theoretically justified superposition in accordance with the *quantum probabilistic description* of molecules and is not subject to *arbitrary* atom−atom correspondences. Approximate models of molecular electron densities permit a rapid evaluation of the similarity integrals. The atomic shell approximation (ASA)[7] is a fitting technique of *ab initio* molecular densities to linear combinations of Gaussian 1*s* functions located at the nuclei. The simplicity inherent to the ASA densities does not impact on the extremely accurate values for the similarity integral (2) as extensively analyzed in previous works.[7,8] The partition of the electron density into atomic contributions is also advantageous for the study of molecular fragments. Furthermore, a global search algorithm for the maximization of the overlap density function can be established, providing safeguards to unbiased and reproducible molecular alignments.

The present contribution defines a practical algorithm derived from the global maximum solution for nuclear collapsed, electron distributions. The extension of this limited global solution to the simplified ASA densities is intended, and simpler, heuristic routines are also proposed. The efficiency and robustness of these algorithms is analyzed.

Toward a Global Maximization

The similarity function $z_{AB}(\mathbf{\Omega})$ presents a considerable number of maxima. The number of maxima, and therefore the complexity of the optimization, increases according to the size of the involved molecules. The appearance and this augmenting in complexity of function $z_{AB}(\mathbf{\Omega})$ is visualized in Figure 1 for the molecular systems N/N_3^-, N_2/N_3^-, and N_3^-/N_3^-, computed at the ASA level. Similar pictures comparing ASA similarity functions with the MP2/6−311G** ones are published elsewhere.[8] Deviations are lesser and only noticeable in a difference plot.

The relevant information of the QMSM function for simple cases consisting of linear molecules is manifested, once the molecules are aligned, by translating one of them along the internuclear axis. The complexity of the whole function $z_{AB}(\mathbf{\Omega})$ can be deduced from these figures when studying a general case with nonlinear and larger molecules. Systematic or stochastic explorations over all six variables $\mathbf{\Omega}$ are not easily feasible in a global maximum search, because the shape of the function would require an extremely finely divided grid or a very large number of random evaluations of $z_{AB}(\mathbf{\Omega})$. This section proposes the methodology to be used in a robust and efficient optimization of the similarity function. First, the ASA fitting algorithm is sketched. Then, the ASA electron densities are collapsed into Dirac delta functions to define a global search algorithm for the resulting similarity function. Next, three levels of a global search approach are derived from this algorithm that are

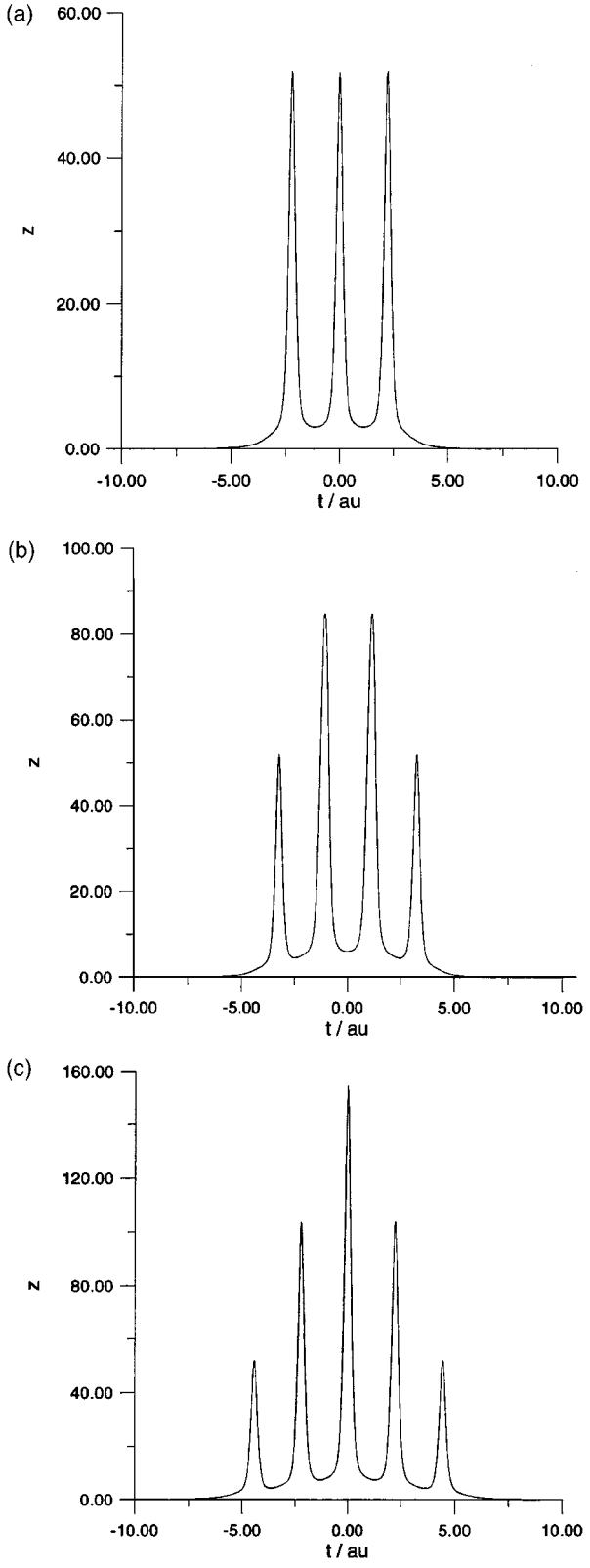


FIGURE 1. ASA similarity functions for linear molecules with respect to the translation along the internuclear axis: (a) N / N_3^-; (b) N_2 / N_3^-; (c) N_3^- / N_3^-.

designed to optimize the original, nondeformed ASA similarity function. To conclude this section, a simple, Lorentzian-like similarity function is introduced for augmenting the computational efficiency and gradient and Hessian formulae are derived at the ASA similarity local optimization level.

ASA AND EMPIRICAL ASA (EASA) ELECTRON DENSITIES

ASA is an algorithm for fitting molecular *ab initio* electron densities to a linear combination of atomic shells,

$$\rho_{\mathrm{ASA}}(\mathbf{r}) \equiv \sum_a \sum_{i \in a} n_i S_i(\mathbf{R}_a - \mathbf{r}), \tag{3}$$

where shells $S_i(\mathbf{R}_a - \mathbf{r})$ are Gaussian $1s$ functions centered at nuclear positions $\mathbf{R}_a$ for each atom a in a given molecule,

$$S_i(\mathbf{R}_a - \mathbf{r}) \equiv \left(\frac{\zeta_i}{\pi}\right)^{3/2} e^{-\zeta_i(\mathbf{R}_a - \mathbf{r})^2}, \tag{4}$$

and the coefficients n_i are the electron population numbers. The physical constraints

$$\sum_i n_i = N, \tag{5}$$

i.e., normalization, and the set of inequalities

$$n_i \geq 0 \quad \forall i, \tag{6}$$

ensuring a positive value for $\rho_{\mathrm{ASA}}(\mathbf{r})$ in its whole domain, are explicitly considered in the ASA method. The ASA fitting is performed by expanding $\rho_{\mathrm{ASA}}(\mathbf{r})$ within a nearly complete functional space and by minimizing the quadratic integral error function,

$$\varepsilon^2(\mathbf{n}, \zeta) = \int (\rho(\mathbf{r}) - \rho_{\mathrm{ASA}}(\mathbf{r}))^2 \, d\mathbf{r}, \tag{7}$$

employing a variational and iterative selection of compatible functions or shells. Even-tempered basis functions[9] are used for the initial nearly complete space to circumvent the exponent optimization problem. In the minimization, the integral error function (7) is regarded as a function of shell populations $\mathbf{n}$ only. The iterative process starts at an arbitrary positive trial point $\mathbf{n}$, which is improved by jumping, along the shortest path to the minimum of the positive definite quadratic form $\varepsilon^2(\mathbf{n})$, to the point where one or more of the $\mathbf{n}$ components became zero. In the next step, a re-

duced functional space is defined by rejecting from the nearly complete space of functions those with zero populations and positive slope, which would contribute with negative populations in an infinitesimal steepest descent displacement, and a new shortest path and improved vector **n** is computed in the reduced space. The process stops when the minimum in a physically compatible subspace is reached and slopes of all discarded functions are positive, the condition of a restricted minimum. At the termination of the ASA process an optimal, physically sound and computationally simple model of molecular electron densities is obtained.

The reduction of computation time that these simplified densities introduce in the similarity evaluation is, as one can infer from its simple form, several orders of magnitude when compared with *ab initio* computations. However, the required *ab initio* densities, previous to the calculation of the ASA ones, might still be an undesirable difficulty. In these cases, empirical models of the molecular electron densities are advantageous. The so-called *promolecular*[10] electron density expresses molecular densities as sums of the contributions of free atoms, in the form

$$\rho_{\mathrm{EASA}} = \sum_a q_a \, \rho^a_{\mathrm{ASA}}(\mathbf{R}_a - \mathbf{r}). \tag{8}$$

The coefficients q_a, if any, here take care of the atomic charge transfers in the molecular environment and contributions $\rho^a_{\mathrm{ASA}}(\mathbf{R}_a - \mathbf{r})$ are free atom densities. To reduce the number of functions in $\rho^a_{\mathrm{ASA}}(\mathbf{R}_a - \mathbf{r})$, a CNDO-like simplification was proposed[1,8] that describes atomic densities as a single squared function,

$$\rho^a_{\mathrm{EASA}} = |S^{n_a}_a(\mathbf{R}_a - \mathbf{r})|^2. \tag{9}$$

In eq. (9), $S^{n_a}_a(\mathbf{R}_a - \mathbf{r})$ are spherical Slater or Gaussian functions of order n_a, with n_a being the row in the Periodic Table of each atom a. The shell exponents $\{\zeta_a\}$ are taken as the ones reproducing the *ab initio* self-similarity measure of the atom. These electron densities were called EASA[8] due to the formal resemblance to the ASA functions and due to its non *ab initio* nature.

The ASA or EASA derived similarity function results in a sum of isotropic atom–atom terms,

$$z_{AB}(\mathbf{\Omega}) = \sum_{a \in A} \sum_{b \in B} z_{ab}(r_{ab}(\mathbf{\Omega})). \tag{10}$$

These simple atom–atom similarity contributions are given by

$$z_{ab}(r_{ab}(\mathbf{\Omega})) = \sum_{i \in a} \sum_{j \in b} n_i n_j s_{ij}(r_{ab}(\mathbf{\Omega})), \tag{11}$$

where s_{ij} is the overlap integral of the shells S_i and S_j.

LIMIT CASE: DEFINING A GLOBAL SEARCH ALGORITHM

In the limit of electron densities infinitely compacted into the nuclei, the set of strictly positive values of the derived similarity function is finite and, therefore, the identification of the maximum element in the set is feasible. Introducing a deforming parameter t in the ASA density expression,

$$\rho_{\mathrm{ASA}}(\mathbf{r}; t) \equiv \sum_a \sum_{i \in a} n_i S_i(\mathbf{R}_a - \mathbf{r}; t), \tag{12}$$

permitting the compression of the atomic shells,

$$S_i(\mathbf{R}_a - \mathbf{r}; t) \equiv \left(\frac{\zeta_i t}{\pi}\right)^{3/2} e^{-\zeta_i t(\mathbf{R}_a - \mathbf{r})^2}, \tag{13}$$

one obtains the collapsed electron density in terms of Dirac delta functions,

$$\tilde{\rho}_A(\mathbf{r}) \equiv \lim_{t \to \infty} \rho_A(\mathbf{r}; t) = \sum_a q_a \delta(\mathbf{R}_a - \mathbf{r}). \tag{14}$$

The population numbers q_a in the collapsed function $\tilde{\rho}_A(\mathbf{r})$ are the sum

$$q_a = \sum_{i \in a} n_i. \tag{15}$$

Introducing these limit densities in eq. (2), the similarity function becomes

$$\tilde{z}_{AB}(\mathbf{\Omega}) = \sum_a \sum_b q_a q_d \delta(\mathbf{R}_a - \mathbf{R}_b(\mathbf{\Omega})). \tag{16}$$

To define a search algorithm, it is proposed to classify all possible nonzero values of $\tilde{z}_{AB}(\mathbf{\Omega})$ into the three following subsets of values. The first one, $\mathbb{Z}_{ab}$, contains the function values arising from the superposition of each atom a of molecule A with each atom b of molecule B. It will be defined as

$$\mathbb{Z}_{ab} \equiv \{q_a q_b \delta(0)\}, \tag{17}$$

and its maximizer $\tilde{z}^*_{ab}$ will be

$$\tilde{z}^*_{ab} \equiv \max(\mathbb{Z}_{ab}). \tag{18}$$

The second subset, $\mathbb{Z}_{aba'b'}$, is the one including the values coming from the simultaneous superposition of two and only two atoms, that is,

$$\mathbb{Z}_{aba'b'} \equiv \{(q_a q_b + q_{a'} q_{b'})\delta(0) \,|\, d_{aa'} = d_{bb'}\}. \quad (19)$$

The existence of elements in $\mathbb{Z}_{aba'b'}$ comes conditioned to the existence of equal interatomic distances in molecules A and B. The maximizer of $\mathbb{Z}_{aba'b'}$ is identified as

$$\tilde{z}^*_{aba'b'} \equiv \max(\mathbb{Z}_{aba'b'}). \quad (20)$$

The last defined subset $\mathbb{Z}_{aba'b'a''b''}$ contains the remaining of all possible nonzero values, being defined as

$$\mathbb{Z}_{aba'b'a''b''} \equiv \{(q_a q_b + q_{a'} q_{b'} + q_{a''} q_{b''})\delta(0) + Q_{aba'b'a''b''} \,|\, d_{aa'} = d_{bb'} \wedge d_{aa''} = d_{bb''} \wedge d_{a'a''} = d_{b'b''}\}. \quad (21)$$

The superposition of three atoms, if possible for the structures A and B, implies the use of all six degrees of freedom $\mathbf{\Omega}$, translations and rotations, of the moving molecule B. Therefore, one can define the subsets $\mathbb{Z}_{ab}$ and $\mathbb{Z}_{aba'b'}$ but not a set of strict coincident three-atom values. The notation $\mathbb{Z}_{aba'b'a''b''}$ should be read as the subset of values obtained when molecule B is translated to overlap atoms b and a, orientated to superimpose b' and a', and finally rotated along the **aa**′ axis until b'' coincides with a''. This univocal defines a point in the function domain $\mathbf{\Omega}$. Then, the term $Q_{aba'b'a''b''}$ is added to include the simultaneous superposition of other atoms, being

$$Q_{aba'b'a''b''} \equiv \sum_{a''' \neq a, a', a''} \sum_{b''' \neq b, b', b''} q_{a'''} q_{b'''} \delta \times (\mathbf{R}_{a'''} - \mathbf{R}_{b'''}(a, b, a', b', a'', b'')). \quad (22)$$

Analogously to the other subsets, the maximizer of $\mathbb{Z}_{aba'b'a''b''}$ is given by

$$\tilde{z}^*_{aba'b'a''b''} \equiv \max(\mathbb{Z}_{aba'b'a''b'}), \quad (23)$$

and the global maximizer of the similarity function $\tilde{z}_{AB}(\mathbf{\Omega})$ finally appears to be symbolically defined as

$$\tilde{z}^*_{AB} = \max(z^*_{ab}, z^*_{aba'b'}, z^*_{aba'b'a''b''}). \quad (24)$$

The above classification suggests the algorithm for the identification of $\tilde{z}^*_{AB}$, which is presented in Table I. A resembling algorithm, related to ligand binding calculations, has also been reported as the *docking problem* algorithm.[11] A subtle detail should be noted at that point. In the case where atoms a, a', a'' and b, b', b'' are lying along a line, and the distance criteria in (21) is accomplished, the values of the similarity function classified in $\mathbb{Z}_{aba'b'}$,

$$\{(q_a q_b + q_{a'} q_{b'})\delta(0), (q_a q_b + q_{a''} q_{b''})\delta(0), (q_{a'} q_{b'} + q_{a''} q_{b''})\delta(0)\}, \quad (25)$$

do not exist for the same reason argued when introducing the subset $\mathbb{Z}_{aba'b'a''b''}$. However, the identification of $\tilde{z}^*_{AB}$ as given in Table I is general for any structure A and B and therefore applicable in this particular case.

The extreme deformation applied to $z_{AB}(\mathbf{\Omega})$ reveals the essence of this function and allows the definition of a global search scheme. The validity of such an approach to scan the nondeformed similarity function is enforced by the dazzling results obtained from its application. Deforming functions for the localization of its global extremes has also been proposed in the context of atom cluster optimization, through the so-called diffusion equation method.[12]

TABLE I.
Global Maximum Identification Algorithm of Deformed Similarity Function $\tilde{z}_{AB}$.

```
INITIALIZE z̃*_AB = 0
DO FOR a ∈ A
DO FOR b ∈ B
  z̃*_AB = max(z̃*_AB, q_a q_b δ(0))
  DO FOR a′ ∈ A \ a
  DO FOR b′ ∈ B \ b
    IF d_aa′ = d_bb′ THEN
      z̃*_AB = max(z̃*_AB, q_a q_b δ(0) + q_a′ q_b′ δ(0))
      DO FOR a″ ∈ A \ a ∧ a′
      DO FOR b″ ∈ B \ b ∧ b′
        IF d_aa″ = d_bb″ ∧ d_a′a″ = d_b′b″ THEN
          TRANSLATE b TO a
          ALIGN bb′ WITH aa′
          ROTATE ALONG bb′ UNTIL d_a″b″ = 0
          z̃*_AB = max(z̃*_AB, z̃_AB(a, b, a′, b′, a″, b″))
        END IF
      END DO FOR b″
      END DO FOR a″
    END IF
  END DO FOR b′
  END DO FOR a′
END DO FOR b
END DO FOR a
GLOBAL MAXIMUM z̃*_AB
```

PRACTICAL QUASIGLOBAL ALGORITHMS FOR ASA/EASA DENSITIES

The scheme for the global maximization of $\tilde{z}_{AB}$ is the basis to derive a set of practical algorithms for nondeformed similarity functions $z_{AB}(\mathbf{\Omega})$. The direct transformation of the global algorithm in Table I to ASA similarities will be constructed as the one in Table II. The equality condition,

$$d_{aa'} = d_{bb'}, \tag{26}$$

that permits the simultaneous superposition of the pairs $a-b$ and $a'-b'$ is translated into the condition

$$z_{a'b'}(\min(d_{a'b'})) > \varepsilon_1, \tag{27}$$

meaning that the orientation of **bb**′ to **aa**′ will only be explored if the contribution of $z_{a'b'}$ in z_{AB} is superior to a threshold value ε_1, instead of the most restrictive condition (26). The distance $d_{a'b'}$, once a and b are superimposed, will be a minimum and therefore $z_{a'b'}$ a maximum when vector **bb**′ is oriented to vector **aa**′, being given by the difference

$$\min(d_{a'b'}) = |d_{aa'} - d_{bb'}|, \tag{28}$$

which is computable just in terms of the constant intermolecular distances.

TABLE II.
Global, Complete Search Algorithm of Third-Level Approach (LA/III).

```
INITIALIZE z*_AB = 0
DO FOR a ∈ A
DO FOR b ∈ B
  TRANSLATE b TO a
  DO FOR a′ ∈ A \ a
  DO FOR b′ ∈ B \ b
    IF z_a′b′(min(d_a′b′)) > ε1  THEN
      ALIGN bb′ WITH aa′
      DO FOR a″ ∈ A \ a ∧ a′
      DO FOR b″ ∈ B \ b ∧ b′
        IF z_a″b″(min(d_a″b″)) > ε2  THEN
          ROTATE ALONG bb′ AT MINIMUM d_a″b″
          z*_AB = max(z*_AB, z_AB(a, b, a′, b′, a″, b″))
        END IF
      END DO FOR b″
      END DO FOR a″
    END IF
  END DO FOR b′
  END DO FOR a′
END DO FOR b
END DO FOR a
GLOBAL MAXIMIZER ESTIMATE z*_AB =
  z_AB(a*, b*, a′*, b′*, a″*, b″*)
```

Similarly, the condition

$$d_{aa''} = d_{bb''} \wedge d_{a'a''} = d_{b'b''} \tag{29}$$

for the simultaneous matching of $a-b$, $a'-b'$, and $a''-b''$ now takes the form

$$z_{a''b''}(\min(d_{a''b''})) > \varepsilon_2. \tag{30}$$

As before, the minimum distance $d_{a''b''}$ can be inferred directly for the set of intermolecular distances, having

$$\min(d_{a''b''}) = \left[(d_a^x - d_b^x)^2 + (d_a^y - d_b^y)^2\right]^{1/2} \tag{31}$$

in terms of d^x and d^y defined, for the A molecule, as

$$d_a^x \equiv \left[d_{aa'}^2 + d_{aa''}^2 - d_{a'a''}^2\right]/2d_{aa'} \tag{32}$$

and

$$d_a^y \equiv \left[d_{aa''}^2 - d_a^{x^2}\right]^{1/2}, \tag{33}$$

respectively. For any set of atoms a, b, a', b', a'', b'' simultaneously accomplishing conditions (27) and (30), a transforming matrix aligning **bb**′ to **aa**′ and rotating along the **bb**′ axis to minimize $d_{a''b''}$ is computed. The coordinate transformation is applied to each atom in molecule B and z_{AB} is finally evaluated. The algorithm in Table I suggests a scanning scheme of function $z_{AB}(\mathbf{\Omega})$ where just the values $z_{AB}(a, b, a', b', a'', b'')$ are considered. One might expect that the maximum value

$$z_{AB}^* = z_{AB}(a^*, b^*, a'^*, b'^* a''^*, b''^*) \tag{34}$$

will be a true estimate of the global maximizer.

If restrictions (27) and (30) were not considered, the identification of the global maximizer z_{AB}^* will be a process requiring

$$n_A n_B (n_A - 1)(n_B - 1)(n_A - 2)(n_B - 2) \approx n_A^3 n_B^3 \tag{35}$$

evaluations of the similarity function $z_{AB}(\mathbf{\Omega})$. Numerical values of the thresholds ε_1 and ε_2 are determined in the next section by computations on a test involving a variety of molecules. Their use in restrictions (27) and (30) make it possible that just a small fraction of the $n_A^3 n_B^3$ evaluations are required. However, the algorithm is still cumbersome at the present technological level and for

TABLE III.
Global, Search Algorithm of Second-Level Approach (LA / II).

INITIALIZE $z^*_{AB} = 0$
DO FOR $a \in A$
DO FOR $b \in B$
  DO FOR $a' > a$
  DO FOR $b' > b$
    IF $z_{a'b'}(\min(d_{a'b'})) > \varepsilon_1$ THEN
      REDEFINE $a, b, a', b' / z_{ab}\ (0) > z_{a'b'}(0)$
      DEFINE a'' AND b'' MAXIMIZING $z_{a''b''}(\min(d_{a''b''}))$, $\forall a'', \forall b''$
      TRANSLATE b TO a
      ALIGN **bb**′ WITH **aa**′
      ROTATE ALONG **bb**′ AT MINIMUM $d_{a''b''}$
      $z^*_{AB} = \max(z^*_{AB}, z_{AB}(a, b, a', b', a'', b''))$
    END IF
  END DO FOR b'
  END DO FOR a'
END DO FOR b
END DO FOR a
QUASIGLOBAL MAXIMIZER ESTIMATE $z^*_{AB} = z_{AB}(a^*, b^*, a'^*, b'^*)$

large molecules. In the following, the two possible decouplings of the three levels of nested loops are considered. The algorithm in Table III introduces the decoupling of the innermost level of loops, evaluating z_{AB} just for the pair a''–b'' with the maximum $z_{a''b''}$ value. The redefinition of the pairs a–b and a'–b' should be noted in Table III and after condition (27). It has proved favorable that the pair defining the molecular translation was the one with the greatest value of the atom–atom similarity function at zero distance. This can be understood as the maximization of the two terms

$$z_{ab} + z_{a'b'} \tag{36}$$

in $z_{AB}(\mathbf{\Omega})$. The efficiency of the algorithm, associated to the number of function evaluations, will be

$$\tfrac{1}{2}n_A(n_A - 1)\tfrac{1}{2}n_B(n_B - 1) \approx \tfrac{1}{4}n_A^2 n_B^2. \tag{37}$$

A further decoupling is also defined, giving the algorithm in Table IV. Now evaluation of z_{AB} is extracted from the loop over the orienting pair a'–b', resulting in a $n_A n_B$ process.

COMPUTATIONAL ACCELERATION: ATOM–ATOM LORENTZIAN SIMILARITY FUNCTIONS

The search for an estimate of the global maximizer requires multiple evaluations of the atom–atom similarity function $z_{ab}(r_{ab})$. This search

TABLE IV.
Simplest Global Search: Algorithm of First-Level Approach (LA / I).

INITIALIZE $z^*_{AB} = 0$
DO FOR $a \in A$
DO FOR $b \in B$
  DEFINE a' AND b' MAXIMIZING $z_{a'b'}(\min(d_{a'b'}))$, $\forall a', \forall b'$
  REDEFINE $a, b, a', b' / z_{ab}(0) > z_{a'b'}(0)$
  DEFINE a'' AND b'' MAXIMIZING $z_{a''b''}(\min(d_{a''b''}))$, $\forall a'', \forall b''$
  TRANSLATE b TO a
  ALIGN **bb**′ WITH **aa**′
  ROTATE ALONG **bb**′ AT MINIMUM $d_{a''b''}$
  $z^*_{AB} = \max(z^*_{AB}, z_{AB}(a, b, a', b', a'', b''))$
END DO FOR b
END DO FOR a
QUASIGLOBAL MAXIMIZER ESTIMATE $z^*_{AB} = z_{AB}(a^*, b^*)$

can be speeded up if $z_{ab}(r_{ab})$ is fitted to a Lorentzian-like function,

$$z_{ab}(r_{ab}) \cong \frac{z_{ab}(0)}{1 + b_{ab}^2 r_{ab}^{c_{ab}}}. \qquad (38)$$

This simpler functional dependency has proven to be flexible enough to fit $z_{ab}(r_{ab})$ to a sufficient accuracy for all atom pairs involving H to Kr. The variational parameters b_{ab}^2 and c_{ab}, which mainly describe the shape of the atomic similarity function, have been determined for free atoms while the $n_A n_B$ values $z_{ab}(0)$, depending on charge transfers and on density basis sets, are computed at the beginning of the maximization process. This gives a function exact at its maximum that decreases approximately, following the dependence of free atoms. Due to the fact that, in the search process, important values are the ones close to the maxima of those atom–atom functions, the least squares fitting is performed from zero to an $r_{\max}$ taken as 0.5 au, and making parameter b_{ab}^2 squared, to ensure that extrapolations to larger distances behave according to the original function. The trial value of the exponent c_{ab} in the nonlinear fitting was taken equal to two, while the initial b_{ab}^2 was chosen as

$$b_0^2 = \frac{1}{0.25^2}\left(\frac{z(0)}{z(0.25)} - 1\right), \qquad (39)$$

to reproduce the value of the function in the middle of the interval. Optimal b_{ab}^2 and c_{ab} parameters were determined using the Levenberg–Marquardt algorithm[13] fitting 100,000 points of the ASA similarity function. The ASA densities were derived from atomic HF/3–21G *ab initio* electron densities. Examples of such parameters are presented in Table V and the corresponding functions are pictured in Figure 2. Note that the function values for r_{ab} above 0.5 au are extrapolated.

TABLE V.
Parameters of Lorentzian-Like Atom – Atom Similarity Function for Carbon–Halogen Pairs.

ab	$z_{ab}(0)$	b_{ab}^2		c_{ab}		$\bar{\varepsilon}$
C–F	57.815	46.712	*0.0041*	2.198	*0.0007*	0.262
C–Cl	142.507	33.667	*0.0013*	1.939	*0.0003*	0.853
C–Br	380.931	45.169	*0.0006*	2.039	*0.0001*	2.822

Uncertainties appear in italics. The values $\bar{\varepsilon}$ are the mean absolute errors of function $z_{ab}(r_{ab})$ in the fitting interval.

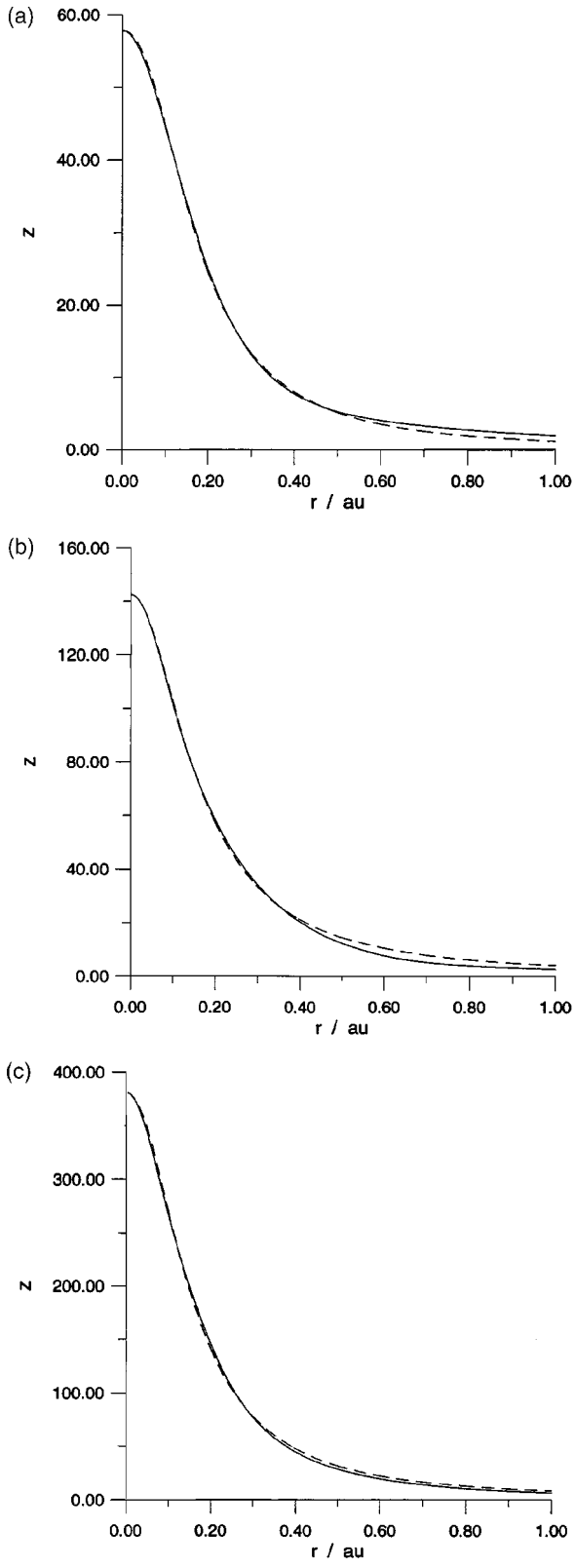


FIGURE 2. Carbon–halogen similarity functions $z_{ab}(r_{ab})$. Solid lines represent the ASA function for free atoms and dashed lines represent the Lorentzian-like functions: (a) z_{C-F}; (b) z_{C-Cl}; (c) z_{C-Br}.

SOLUTION REFINING: EFFICIENT IMPLEMENTATION OF NEWTON ALGORITHM

Once an estimate of a maximizer is identified, a usual, local optimization will terminate the similarity measure process. The analytical gradient and Hessian matrix are computed by summing the corresponding derivatives of the two shell terms in the similarity function. These terms take the form

$$z_{ij}(\mathbf{\Omega}) = n_i n_j c_i c_j \left(\frac{\pi}{\zeta_i + \zeta_j}\right)^{3/2} \times \exp\left[-\frac{\zeta_i \zeta_j}{\zeta_i + \zeta_j} r_{ab}^2(\mathbf{\Omega})\right], \quad i \in a, \quad j \in b, \tag{40}$$

where c_i and c_j are the constants of normalization. Vector $\mathbf{\Omega}$ represents the set of translations and rotation angles indicating the spatial positioning of molecule B, with respect to a set of Cartesian coordinates where the two molecules A and B are placed. One has

$$\mathbf{\Omega} \equiv (t_x, t_y, t_z, r_x, r_y, r_z). \tag{41}$$

The explicit form for the distance between atoms a and b, $r_{ab}(\mathbf{\Omega})$, appears as

$$r_{ab}(\mathbf{\Omega}) = \left|\mathbf{r}_a - \left(\mathbf{R}(r_x, r_y, r_z).\mathbf{r}_b + \mathbf{T}(t_x, t_y, t_z)\right)\right|, \tag{42}$$

in terms of the rotation matrix $\mathbf{R}(r_x, r_y, r_z)$ and the translation vector $\mathbf{T}(t_x, t_y, t_z)$. In the particular case of $\mathbf{\Omega} = \mathbf{0}$, where no transformation has been applied to molecule B, the gradient and the Hessian are determined by the following simple formulae. The first component of the gradient is

$$g^0_{t_x} \equiv \left.\frac{\partial z_{ij}}{\partial t_x}\right|_{\mathbf{\Omega}=\mathbf{0}} = 2 z_{ij} \frac{\zeta_i \zeta_j}{\zeta_i + \zeta_j}(x_a - x_b), \tag{43}$$

with analogous expressions for the two other terms involving translations. The component of the gradient containing the rotation along the x axis is

$$g^0_{r_x} \equiv \left.\frac{\partial z_{ij}}{\partial r_x}\right|_{\mathbf{\Omega}=\mathbf{0}} = 2 z_{ij} \frac{\zeta_i \zeta_j}{\zeta_i + \zeta_j}(z_a y_b - y_a z_b). \tag{44}$$

The two remaining rotation components are correspondingly defined. The Hessian can be expressed in terms of the function itself and the gradient components. For the diagonal terms it is obtained by

$$h^0_{t_x t_x} \equiv \left.\frac{\partial^2 z_{ij}}{\partial t_x^2}\right|_{\mathbf{\Omega}=\mathbf{0}} = \frac{g^0_{t_x} g^0_{t_x}}{z_{ij}} - 2 z_{ij} \frac{\zeta_i \zeta_j}{\zeta_i + \zeta_j}, \tag{45}$$

for the translations. For the rotations, the following expression applies:

$$h^0_{r_x r_x} \equiv \left.\frac{\partial^2 z_{ij}}{\partial r_x^2}\right|_{\mathbf{\Omega}=\mathbf{0}} = \frac{g^0_{r_x} g^0_{r_x}}{z_{ij}} - 2 z_{ij} \frac{\zeta_i \zeta_j}{\zeta_i + \zeta_j}(y_a y_b - z_a z_b). \tag{46}$$

As before, the rest of the diagonal terms are analogously defined. The nondiagonal terms $h^0_{t_x t_y}$, $h^0_{t_x t_z}$, $h^0_{t_x r_x}$, $h^0_{t_y t_z}$, $h^0_{t_y r_y}$, and $h^0_{t_z r_z}$ have the form

$$h^0_{t_x t_y} \equiv \left.\frac{\partial^2 z_{ij}}{\partial t_x\, \partial t_y}\right|_{\mathbf{\Omega}=\mathbf{0}} = \frac{g^0_{t_x} g^0_{t_y}}{z_{ij}}. \tag{47}$$

The rest of the terms involving rotations and translations are

$$h^0_{t_x r_y} \equiv \left.\frac{\partial^2 z_{ij}}{\partial t_x\, \partial r_y}\right|_{\mathbf{\Omega}=\mathbf{0}} = \frac{g^0_{t_x} g^0_{r_y}}{z_{ij}} - 2 z_{ij} \frac{\zeta_i \zeta_j}{\zeta_i + \zeta_j} z_b, \tag{48}$$

also being similar to $h^0_{t_y r_z}$ and $h^0_{t_z r_y}$. The terms $h^0_{t_x r_z}$, $h^0_{t_y r_x}$, and $h^0_{t_z r_y}$ change the minus by a plus sign in (48). And finally, the remaining elements of the Hessian involving just rotations have the form

$$h^0_{r_x r_y} \equiv \left.\frac{\partial^2 z_{ij}}{\partial r_x\, \partial r_y}\right|_{\mathbf{\Omega}=\mathbf{0}} = \frac{g^0_{r_x} g^0_{r_y}}{z_{ij}} + 2 z_{ij} \frac{\zeta_i \zeta_j}{\zeta_i + \zeta_j} x_a y_b, \tag{49}$$

and equivalent expressions for $h^0_{r_x r_z}$ and $h^0_{r_y r_z}$. The computation of the Hessian eigenvalues after the global search showed that, in most of the test cases, the maximizer lies in the basin of a maximum stationary point. Therefore, a simple iterative Newton process will be, in general, successful. This give the improved vector $\mathbf{\Omega}^1$ as

$$\mathbf{\Omega}^1 = -\mathbf{H}(\mathbf{0})^{-1}\mathbf{g}(\mathbf{0}). \tag{50}$$

Note that $\mathbf{\Omega}^0$ does not appear in eq. (50) because it was taken as a null vector. Once $\mathbf{\Omega}^1$ has been determined, one can compute an approximate value of $z_{AB}(\mathbf{\Omega}^1)$ in the assumption of a quadratic form. In the case where this approximate $z_{AB}(\mathbf{\Omega}^1)$ is smaller in value than the previous exact computation, $z_{AB}(\mathbf{\Omega}^0)$, one can suspect, circumventing

the eigenvalue computation, that the point $\mathbf{\Omega}^1$ is not in the basin of a maximum. In such a case, redefining $\mathbf{\Omega}^1$ as

$$\mathbf{\Omega}^1 = \mathbf{H}(\mathbf{0})^{-1}\mathbf{g}(\mathbf{0}) \tag{51}$$

proved to be sufficient for safeguarding the Newton's method.[14] At this point of the process, the rotation matrix $\mathbf{R}(\mathbf{\Omega}^1)$ and the translation vector $\mathbf{T}(\mathbf{\Omega}^1)$ are evaluated and the coordinate transformation is applied to molecule B. Note that if one refers $\mathbf{\Omega}$ to the new coordinates of B, instead of the original ones, the simpler formulae (43)–(49) can still be used. This convention circumvents sinusoidal and cosinusoidal evaluations in the calculation of the derivatives. The adopted convergence criterion stops the iterative process when the change in $z_{AB}(\mathbf{\Omega})$ falls below the tolerance value of 1×10^{-8}. In all tested cases, convergence required no more than five cycles. The stationary points found always resulted to be true maxima, as the Hessian diagonolization proved.

Convergence and Efficiency of Global Search

In the previous section, we presented three approximate, global search algorithms derived from the exact solution for hypothetical quantum systems with electron distributions collapsed into the nuclei. The validity of such approximations or *truncations*, as well as the numerical values for the parameters ε_1 and ε_2 introduced earlier, are evaluated on a selection of 18 *real* molecules. This selection embraces a variety of organic molecules that are comparable in size to most of the usual bioactive compounds, including heteroatoms up to the fourth period, and presents similar but also notably different, spatial structures. Chemical formulae and Cambridge Structural Database[15] reference codes are listed in Table VI. Experimental, accessible structures were preferred to elude the multiple conformer problem, and, therefore, to facilitate the reproducibility of the present similarity measures.

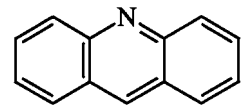
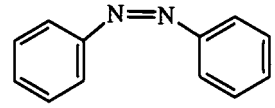
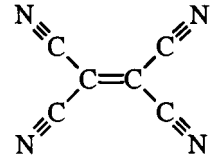
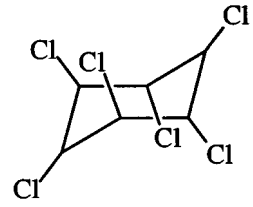
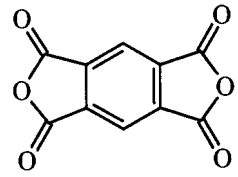
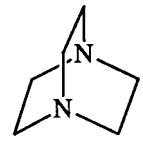
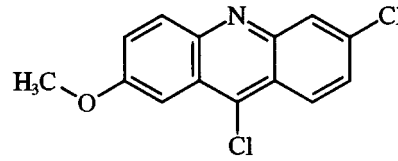
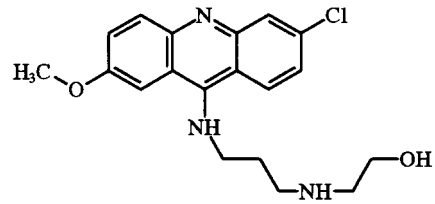
Electron densities for the molecules in Table VI, required for the similarity function determination, were obtained in the following way. First, *ab initio*, HF/3–21G, electron densities were computed at the crystal geometry using the Gaussian 94 ensemble of programs.[16] Then, ASA densities were derived using the ASAC program.[17] Even tempered basis sets given in ref. 7 were considered for the initial, nearly complete space, taking 10, 20, 25, and 30 Gaussian $1s$ functions for each element in the first, second, third, and fourth period, respectively. EASA densities were constructed taking a single Slater $n_a s$ shell for each atom a in the molecule, with the exponents reproducing atomic HF/3–21G self-similarities. Coefficients q_a in the linear combination of atomic densities, in eq. (8), were fixed to unity.

The quality of these approximated electron densities, as well as the one of Lorentzian-like similarity functions, is tested in Table VII by comparison with the *ab initio* self-similarity values. ASA densities provide accurate enough values, not only in self-similarities, but also in all the similarity function domain.[7,8] In Table VII the number of compatible functions or shells for each molecular density is also reported. EASA densities, due to their definition and construction, yield acceptable self-similarity values but somehow erroneous cross-similarity global maxima. Contrarily, the Lorentzian-like approach, as described previously and indicated by L + ASA in the table, exhibits a substantial overestimation, but it is uniform at any point in the function domain as we will see in the global search results. This approach appears to be suitable, therefore, to scan the similarity function. Also in Table VII (denoted by L) we present the self-similarity values with the quantities $z_{ab}(0)$ in function (38) taken as if atoms a and b were isolated, instead of the molecular ASA quantities. This is so for a possible accuracy estimation with the example presented in the next section.

The similarity function optimization was performed excluding in all global searches the hydrogen atoms as translating–orienting atoms, because they do not show differentiated peaks in the similarity function.[8] Global search and local optimization algorithms were implemented, as described in the previous section, in a computer program named ASASim. Speeding techniques, such as parallelization, integer arithmetic in distance evaluation, or dynamic thresholding in integral and derivative computations, have yet to be considered and will be described elsewhere.

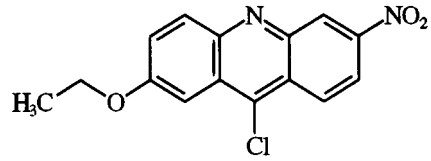
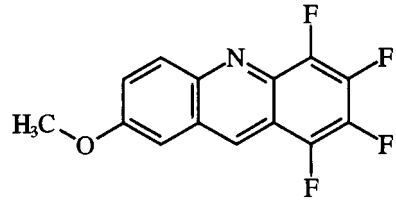
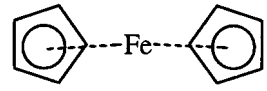
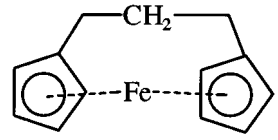
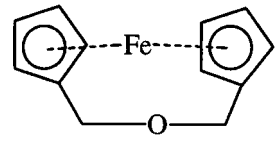
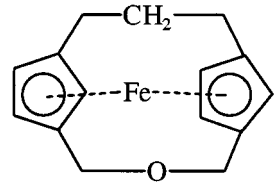
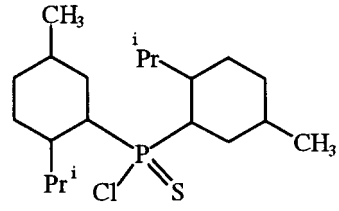
First of all, we performed extensive similarity computations in a subset of the molecules in Table VI, labeled as the group A, to evaluate the two truncations introduced when the algorithm LA/III in Table II was derived from its limit solution analog in Table I. This permitted an estimation of thresholds ε_1 and ε_2 and a verification of the implicit assumption that the best estimate of a maximizer converges to the best, global maxi-

TABLE VI.
Cambridge Structural Data Base Reference Codes, Chemical Formulae, and Structures of Selected Molecules.

Molecule	CSD Ref. Code[15]	Formula	Structure	n_A
A_1	ACRDIN01	$C_{13}H_9N$		14
A_2	AZBENC01	$C_{12}H_{10}N_2$		14
A_3	TCYETY	C_6N_4		10
A_4	HCCYHB	$C_6H_6Cl_6$		12
A_5	PYMDAN	$C_{10}H_2O_6$		16
A_6	TETDAM03	$C_6H_{12}N_2$		8
B_1	BAWMIV	$C_{14}H_9Cl_2NO$		18
B_2	CMACPE	$C_{19}H_{22}ClN_3O_2$		25

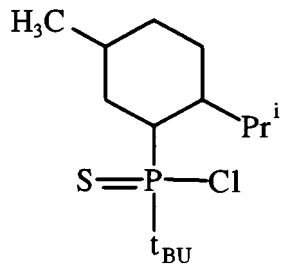
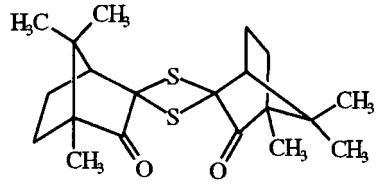
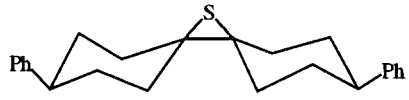
(*Continues on next page*)

TABLE VI.
(Continued)

Molecule	CSD Ref. Code[15]	Formula	Structure	n_A
B_3	HABGIA	$C_{15}H_{11}ClN_2O_3$		21
B_4	WERXIA	$C_{14}H_7F_4NO$		20
C_1	FEROCE12	$C_{10}H_{10}Fe$		11
C_2	FICJUW	$C_{13}H_{14}Fe$		14
C_3	FICKAD	$C_{12}H_{12}FeO$		14
C_4	FICLAE	$C_{15}H_{16}FeO$		17
D_1	VEKREI	$C_{20}H_{38}ClPS$		23

(*Continues on next page*)

TABLE VI.
(Continued)

Molecule	CSD Ref. Code[15]	Formula	Structure	n_A
D_2	DUPDAT	$C_{14}H_{28}ClPS$		17
D_3	YONROI	$C_{20}H_{28}O_2S_2$		24
D_4	LAPZEH	$C_{24}H_{28}S$		25

The number of nonhydrogen atoms is n_A.

mizer. Moreover, this subset of molecules provided a comparison of the present maximization technique with the usual Monte Carlo method.

Suitable values for thresholds ε_1 and ε_2 will be the most discriminating ones that ensure that no best estimate of the maximizer is missing. The threshold ε_1 was estimated according to the second-level algorithm, LA/II, by means of successive similarity measure computations among the group A molecules. Starting from a value of zero, parameter ε_1 was increased until reaching the highest value below which all best maximizers are considered. Once ε_1 is determind, one can estimate ε_2 in an analogous way, using the third-level algorithm, LA/III. The obtained values for ε_1 are 14, 2, and 25 au for the ASA, Lorentzian-like, and EASA approximations of the similarity function, respectively, while the values for threshold ε_2 are 0.05, 0.15, and 0.4 au for the same sequence of approximate similarity functions.

The reduction in the number of required function evaluations in respect the absence of the ε_1 and ε_2 thresholds can be appreciated in Table VIII.

Local optimizations at each scanned point were performed at the full ASA level for the molecules in group A. The results, together with the ones obtained optimizing only the best estimate of a maximizer, are presented in Table IX. Optimization of the best estimate, if it comes from the scanning scheme in Table II, converges to the global maximum in most of the cases investigated. The slight differences, of less than 0.5 au, appearing in some of the similarity measures are caused by the imprecision of this second truncation to discern among nearly degenerate global maxima. The group A molecules present such split maxima due to a break in their symmetry that was induced by the crystal polarization. More significant differences, corresponding to chemically nonequivalent superposition, occur when the best estimates come from the incomplete scanning schemes. Then, the search algorithm in Table III was trapped in a local maximum in the comparison of molecules A_1 and

TABLE VII.
Performance in Self-Similarity Accuracy of Different Approaches Used.

	z_{HF}	z_{ASA}	n_{ASA}	z_{EASA}	z_{L+ASA}	z_{L}
A_1	457.757	457.623	*162*	460.614	461.230	465.944
A_2	478.720	478.546	*162*	481.185	482.548	486.851
A_3	395.196	395.053	*82*	395.491	395.270	398.017
A_4	6098.443	6098.180	*138*	6101.034	6131.973	6134.756
A_5	793.302	793.084	*162*	795.023	796.540	801.266
A_6	291.532	291.450	*128*	293.235	295.920	297.483
B_1	2539.859	2539.602	*195*	2543.242	2551.499	2556.293
B_2	1897.146	1896.809	*315*	1901.147	1914.616	1916.980
B_3	1798.968	1798.649	*229*	1801.816	1809.932	1814.402
B_4	1041.243	1040.890	*206*	1044.196	1047.699	1053.466
C_1	4259.288	4258.337	*152*	4291.539	4268.684	4272.604
C_2	4354.599	4353.643	*187*	4356.524	4368.411	4370.948
C_3	4402.135	4401.141	*176*	4405.001	4415.495	4419.124
C_4	4496.229	4495.237	*218*	4499.670	4513.431	4517.137
D_1	3070.869	3070.663	*389*	3074.569	3111.479	3108.202
D_2	2881.706	2881.536	*293*	2884.769	2912.934	2911.320
D_3	2400.002	2399.764	*339*	2403.651	2430.479	2430.595
D_4	1558.884	1558.615	*355*	1562.892	1578.797	1580.416

HF indicates the *ab initio* values and ASA and EASA the ones produced by the approximated electron densities, respectively. L + ASA refers to the Lorentzian-like similarity approach corrected by the atomic ASA values at zero distances, while L indicates the pure, promolecular approach. The number of selected functions or shells in the ASA electron densities is denoted by n_{ASA}.

TABLE VIII.
Required Similarity Evaluations for Different Global Search and Similarity Approaches.

	LA / I	LA / II				LA / III			
	N	ASA	L + ASA	EASA	*N*	ASA	L + ASA	EASA	*N*
A_1–A_1	196	1063	1263	1117	*8281*	146,098	50,346	38,056	*4,769,856*
A_1–A_2	196	871	1251	915	*8281*	106,720	57,326	27,038	*4,769,856*
A_1–A_3	140	311	700	381	*4095*	33,106	17,570	7884	*1,572,480*
A_1–A_4	168	306	1059	384	*6006*	121,434	178,176	13,914	*2,882,880*
A_1–A_5	224	765	1559	984	*10,920*	114,502	89,276	32,336	*7,338,240*
A_1–A_6	112	437	505	466	*2548*	46,710	16,758	12,954	*733,824*
A_2–A_2	196	803	1229	847	*8281*	108,000	71,020	29,136	*4,769,856*
A_2–A_3	140	287	637	327	*4095*	27,516	18,692	6848	*1,572,480*
A_2–A_4	168	297	1059	366	*6006*	102,696	201,876	13,020	*2,882,880*
A_2–A_5	224	797	1554	932	*10,920*	117,860	93,424	28,764	*7,338,240*
A_2–A_6	112	352	484	386	*2548*	33,348	16,956	9744	733,824
A_3–A_3	100	281	433	289	*2025*	21,592	10,984	6048	*518,400*
A_3–A_4	120	180	528	258	*2970*	44,328	71,484	4824	*950,400*
A_3–A_5	160	336	894	472	*5400*	39,556	36,252	10,564	*2,419,200*
A_3–A_6	80	126	213	154	*1260*	11,748	4284	2184	*241,920*
A_4–A_4	144	522	1098	558	*4356*	250,488	365,040	25,912	*1,742,400*
A_4–A_5	192	393	1050	435	*7920*	143,168	258,348	15,588	*4,435,200*
A_4–A_6	96	138	510	198	*1848*	38,052	53,136	7848	*443,520*
A_5–A_5	256	1134	2122	1134	*14,400*	203,888	153,184	50,440	*11,289,600*
A_5–A_6	128	312	613	446	*3360*	36,948	25,236	10,800	*1,128,960*
A_6–A_6	64	274	298	286	*784*	33,408	16,366	11,232	*112,896*

LA / I, LA / II, and LA / III refer to the algorithm of the first-, second-, and third-level approaches, respectively. ASA, L + ASA, and EASA indicate the kind of similarity computation. Numbers, *N*, in italics are the similarity evaluations omitting the threshold parameters ε_1 and ε_2.

TABLE IX.
Analysis of Global Maximization Performance in Subset *A* of Molecules.

	ASA						L + ASA		
	LA / I	∇	LA / II	∇	LA / III	∇	LA / I	LA / II	LA / III
A_1–A_1	457.623	457.623	457.623	457.623	457.623	457.623	457.623	457.623	457.623
A_1–A_2	204.525	204.525	204.525	204.525	204.525	204.525	204.525	204.525	204.525
A_1–A_3	*169.091*	170.239	*170.201*	170.239	*170.201*	170.239	*169.091*	*170.201*	*170.201*
A_1–A_4	*224.173*	*231.085*	*224.173*	*231.085*	236.453	236.453	236.453	236.453	236.453
A_1–A_5	198.711	198.711	*198.619*	198.711	*198.619*	198.711	198.711	*198.619*	*198.619*
A_1–A_6	117.911	117.911	117.911	117.911	117.911	117.911	117.911	*117.907*	117.911
A_2–A_2	478.546	478.546	478.546	478.546	478.546	478.546	478.546	478.546	478.546
A_2–A_3	*141.988*	144.562	*144.316*	144.562	*144.316*	144.562	*141.988*	*144.316*	*144.316*
A_2–A_4	292.388	292.388	292.388	292.388	292.388	292.388	*292.076*	*292.076*	*292.077*
A_2–A_5	*164.941*	165.647	165.647	165.647	*165.237*	165.647	*164.941*	*165.237*	*165.237*
A_2–A_6	113.749	113.749	113.749	113.749	113.749	113.749	113.749	113.749	113.749
A_3–A_3	395.053	395.053	395.053	395.053	395.053	395.053	395.053	395.053	395.053
A_3–A_4	*234.728*	238.674	*234.728*	238.674	238.674	238.674	*234.728*	*234.728*	238.674
A_3–A_5	144.507	144.507	144.507	144.507	144.507	144.507	144.507	144.507	144.507
A_3–A_6	*100.452*	*100.979*	*100.452*	*100.979*	*100.980*	100.981	*100.443*	*100.443*	*100.980*
A_4–A_4	6098.180	6098.180	6098.180	6098.180	6098.180	6098.180	6098.180	6098.180	6098.180
A_4–A_5	337.996	337.996	337.996	337.996	337.996	337.996	337.996	337.996	337.996
A_4–A_6	*211.792*	*211.792*	*211.792*	*211.792*	211.802	211.802	211.802	211.802	211.802
A_5–A_5	793.084	793.084	793.084	793.084	793.084	793.084	793.084	793.084	793.084
A_5–A_6	*108.738*	*108.738*	*108.745*	108.747	108.745	108.745	*108.738*	*108.744*	108.745
A_6–A_6	291.450	291.450	291.450	291.450	291.450	291.450	291.450	291.450	291.450

LA / I, LA / II, and LA / III refer to the algorithm of the first-, second-, and third-level approaches, respectively. The ∇ symbol indicates that local optimizations were performed at each estimate of a maximizer. Similarities are computed at the ASA level. L + ASA denotes that the similarity function was scanned using the Lorentzian-like approximation. Roman letters indicate global maxima and italics indicate lower, local maxima.

A_4, while the algorithm of the first level approach, described in Table IV, failed in the cases A_1–A_4, A_2–A_3, and A_3–A_4. Local optimizations at each estimate of a maximizer are between approximately two and three orders of magnitude more time consuming than any of the scanning schemes. Fortunately, the second truncation effect is minor, especially in a complete, third-level search.

Table IX also presents the results obtained after optimizing the best estimates that the rapid, Lorentzian-like similarity functions proportioned. Scanning these Lorentzian-like functions appears to be as accurate as a full ASA maximization, indicating that the Lorentzian overestimation is nearly constant in all function domains. Slight differences leading to other maxima do not result in chemically nonequivalent superposition. By chance, in the measure A_1–A_4, the different, best estimate provided by the Lorentzian-like approach also leads to the global maximum in the two, most approximated, search schemes.

The convergence of the proposed maximization algorithms, in the sense of efficiency, as applied to the well-established, local methods, is tested by the comparison of similarity measures in Table X with the required function evaluations given in Table VIII. They were computed at the three levels of search using the naive, EASA electron densities. In addition a hybrid Monte Carlo and simplex[18] technique was also considered. In this technique, the initial translation is imposed by adopting the same point for the two centers of charge, while rotation angles are randomly established. From this initial move, a simplex maximization was performed. The simplex method was preferred due to its ability to explore the function out of the basin of a given maximizer. Subsequent Monte Carlo moves, inscribed in a box fitting the static molecule, were followed by a simplex maximization only when they were accepted. Once the number of calls to the similarity function reach the ones required in an EASA search, a simplex maximizes the last move and the best measure found is reported. One can appreciate from the results in Table X the quick convergence to global maxima of the proposed three levels of search when com-

TABLE X.
Analysis of Global Maximization Performance in Subset *A* of Molecules.

	LA / I	MC	LA / II	MC	LA / III	MC
A_1–A_1	460.614	*444.981*	460.614	*444.981*	460.614	*447.283*
A_1–A_2	233.966	*80.647*	233.966	*80.647*	233.966	*217.624*
A_1–A_3	191.774	*31.440*	*191.757*	*60.807*	*191.757*	*173.491*
A_1–A_4	302.411	*48.807*	302.411	*223.328*	302.411	*300.608*
A_1–A_5	*242.885*	*242.144*	*242.904*	*242.144*	242.935	*242.144*
A_1–A_6	*122.139*	*114.129*	*124.727*	*114.129*	125.523	*122.902*
A_2–A_2	481.185	*64.879*	481.185	*137.419*	481.185	*379.491*
A_2–A_3	161.887	*43.097*	161.887	*64.462*	161.887	*97.611*
A_2–A_4	328.582	*192.989*	328.582	*192.989*	328.582	*291.726*
A_2–A_5	*199.417*	*71.490*	*209.411*	*119.662*	209.864	*208.907*
A_2–A_6	*114.961*	*47.201*	*119.100*	*60.986*	127.001	*97.066*
A_3–A_3	395.491	*380.313*	395.491	*380.313*	395.491	395.491
A_3–A_4	335.508	*210.824*	335.508	*210.824*	335.508	*293.826*
A_3–A_5	*191.628*	*98.239*	*191.628*	*188.270*	191.704	*188.286*
A_3–A_6	*103.130*	*37.973*	*103.130*	*37.973*	114.157	*92.811*
A_4–A_4	6101.034	*2096.643*	6101.034	*2096.643*	6101.034	*5636.023*
A_4–A_5	503.552	*280.217*	503.552	*280.217*	503.552	*466.635*
A_4–A_6	*239.658*	*95.424*	*239.935*	*95.424*	239.945	*225.516*
A_5–A_5	795.023	*152.433*	795.023	*773.317*	795.023	*794.380*
A_5–A_6	*145.298*	*80.789*	*150.491*	*104.549*	150.813	*143.959*
A_6–A_6	293.235	*93.044*	293.235	*96.606*	293.235	*292.895*

LA / I, LA / II, and LA / III refer to the algorithm of first-, second-, and third-level approaches, respectively. Similarities are computed at the EASA level. The quantities in columns MC are the maxima of the function identified by the hybrid Monte Carlo–simplex technique explained in the text. Roman letters indicate the best maxima found, and italics denote other, lower maxima.

pared to this stochastic technique. In addition, efficiency of the incomplete first- and second-level approach is also manifested.

The results involving the molecules in group A permitted studying the effect on the convergence of the several possible approaches described. These approaches are required to translate the limit and ideal global maximum solution into a computationally efficient algorithm. The remaining molecules provided the test for the computational timing efficiency. Four of these molecules in group B are characterized by the acridine pattern. The four molecules in group C consist of a set of ferrocene compounds. Group D embraces quite dissimilar spatial structures. Similarity measure results are summarized in Table XI.

Computation timing on an IBM RISC 6000/355 workstation is indicated for each pair of molecules and level approach. It includes the scanning of the similarity function through Lorentzian-like functions and the final local maximization performed using ASA densities. Even when the program ASASim is not fully optimized, the first- and second-level approaches are fast enough and accurate for MQSM studies. Those maxima are comparable to the ones in the third-level approach, except for the four cases, B_1–D_2, B_1–D_4, B_2–D_4, and B_4–D_3 that correspond to nonchemically equivalent superposition. Thus, disregarding the nearly degeneracy, and from the complete computations presented in Table IX, one could qualify them as global maxima.

Some of the maxima in Table XI that are identified by the algorithm of the first level are higher than the ones identified by the more complete searches. This *incoherence* is not produced by an overestimation of the threshold parameters, but by the exclusion of local optimization at each estimate. Then, in the most noticeable case, C_2–C_3, the sequence of best estimate values from the simplest search to the complete one is 4084.4, 4192.2, and 4204.0, in accordance with what was expected. The superposition of molecules C_2 and C_3 is nearly equivalent in the two maxima, matching in each case the ferrocene pattern. However, one of the two possible symmetric superpositions of ferrocene also favors the overlap of the oxygen atom in molecule C_3 with the second, methylene carbon

TABLE XI.
ASA Similarity Maxima for Subsets *B*, *C*, and *D* of Molecules.

	LA / I	*t*	LA / II	*t*	LA / III	*t*
B_1–B_2	*1100.172*	*26*	1290.874	*68*	1290.874	*5357*
B_1–B_3	1463.068	*10*	1463.068	*38*	1463.068	*2723*
B_1–B_4	519.381	*8*	519.381	*34*	519.381	*2053*
B_2–B_3	1160.692	*19*	1160.692	*100*	1160.692	*7272*
B_2–B_4	436.171	*15*	436.171	*84*	436.171	*4909*
B_3–B_4	522.292	*12*	522.292	*49*	522.292	*2715*
C_1–C_2	*4014.308*	*6*	*4089.441*	*11*	4089.954	*816*
C_1–C_3	*4018.167*	*3*	4091.414	*10*	4091.414	*742*
C_1–C_4	*4014.352*	*6*	*4073.027*	*13*	4076.034	*1267*
C_2–C_3	4207.126	*12*	*4204.634*	*20*	*4204.634*	*1811*
C_2–C_4	*4055.521*	*8*	*4135.066*	*28*	4135.513	*3311*
C_3–C_4	*4049.278*	*7*	4196.353	*25*	4196.353	*3255*
D_1–D_2	2181.869	*25*	2181.869	*104*	2181.869	*23207*
D_1–D_3	1131.807	*43*	1131.807	*212*	1131.807	*47096*
D_1–D_4	*932.030*	*52*	935.386	*185*	935.386	*22661*
D_2–D_3	1132.127	*17*	1132.127	*91*	1132.127	*19558*
D_2–D_4	*931.659*	*28*	938.600	*83*	938.600	*8990*
D_3–D_4	*871.111*	*25*	881.383	*184*	881.383	*22474*
B_1–C_1	1903.387	*43*	1903.387	*47*	1903.387	*511*
B_1–C_2	1936.081	*9*	1936.081	*19*	1936.081	*1341*
B_1–C_3	*1933.945*	*12*	*1933.945*	*23*	1934.020	*1297*
B_1–C_4	1935.786	*17*	1935.786	*37*	1935.786	*2704*
B_2–C_1	1902.928	*42*	1902.928	*51*	1902.928	*943*
B_2–C_2	*1904.912*	*136*	1905.011	*46*	1905.011	*2649*
B_2–C_3	1905.504	*87*	*1905.473*	*60*	*1905.473*	*2563*
B_2–C_4	1906.876	*25*	1906.876	*74*	1906.876	*5558*
B_3–C_1	1903.013	*80*	1903.006	*79*	1903.006	*676*
B_3–C_2	1935.565	*13*	1935.565	*29*	1935.565	*1698*
B_3–C_3	*1933.867*	*22*	*1933.867*	*37*	1933.980	*1640*
B_3–C_4	1935.770	*29*	*1935.664*	*51*	1935.770	*3416*
B_4–C_1	576.208	*12*	*576.173*	*15*	*576.173*	*547*
B_4–C_2	579.651	*8*	579.651	*23*	579.651	*1472*
B_4–C_3	580.177	*7*	*580.169*	*21*	*580.169*	*1397*
B_4–C_4	584.759	*13*	584.759	*39*	584.759	*2862*
B_1–D_1	1050.327	*13*	1050.327	*70*	1050.327	*10337*
B_1–D_2	*1024.481*	*40*	*1029.296*	*36*	1032.362	*4351*
B_1–D_3	*996.313*	*14*	1011.424	*70*	1011.424	*9360*
B_1–D_4	*975.819*	*15*	*975.819*	*68*	982.596	*5287*
B_2–D_1	1025.439	*24*	1025.439	*194*	1025.439	*31602*
B_2–D_2	*1024.217*	*69*	1027.929	*96*	1027.929	*11863*
B_2–D_3	1018.758	*35*	1018.758	*196*	1018.758	*25150*
B_2–D_4	*924.733*	*41*	*924.733*	*187*	927.365	*14058*
B_3–D_1	1057.650	*16*	1057.650	*100*	1057.650	*13700*
B_3–D_2	*1025.185*	*56*	1039.742	*49*	1039.742	*5646*
B_3–D_3	*1017.975*	*15*	1034.296	*100*	1034.296	*12614*
B_3–D_4	942.362	*33*	942.362	*110*	942.362	*7300*
B_4–D_1	374.533	*22*	374.533	*93*	*374.109*	*10508*
B_4–D_2	369.308	*17*	369.308	*47*	*368.674*	*4471*
B_4–D_3	*568.382*	*14*	*568.382*	*90*	576.718	*9998*
B_4–D_4	353.754	*14*	353.754	*78*	353.754	*5257*
C_1–D_1	*2017.957*	*30*	2018.084	*39*	2018.153	*1829*
C_1–D_2	2011.205	*12*	2011.205	*18*	2011.205	*958*
C_1–D_3	1711.225	*15*	1711.225	*27*	1711.225	*2251*
C_1–D_4	1712.972	*13*	*1712.968*	*21*	*1712.968*	*899*

(*continues on next page*)

TABLE XI.
(Continued)

	LA / I	*t*	LA / II	*t*	LA / III	*t*
C_2–D_1	2032.102	*15*	2032.102	*42*	2032.102	*5582*
C_2–D_2	*2018.101*	*7*	*2018.101*	*20*	2018.156	*2909*
C_2–D_3	*1758.190*	*10*	*1758.190*	*43*	1758.607	*6542*
C_2–D_4	1720.121	*16*	1720.121	*38*	1720.121	*2522*
C_3–D_1	*2028.212*	*11*	2028.967	*37*	2028.967	*5568*
C_3–D_2	2017.576	*17*	2017.576	*31*	2017.576	*2846*
C_3–D_3	1783.307	*10*	1783.307	*42*	1783.307	*6426*
C_3–D_4	*1718.024*	*20*	1718.780	*38*	1718.780	*2548*
C_4–D_1	2029.171	*20*	2029.171	*77*	2029.171	*13018*
C_4–D_2	2018.878	*11*	2018.878	*38*	2018.878	*6392*
C_4–D_3	*1764.400*	*13*	1764.984	*76*	1764.984	*13650*
C_4–D_4	*1725.412*	*21*	1728.708	*67*	1728.708	*5358*

LA / I, LA / II, and LA / III refer to the algorithm of first-, second-, and thid-level approaches, respectively, used in the global search. Computational timing *t*, in seconds on an IBM RISC 6000, includes the Lorentzian-like approach scanning and the final ASA optimization. Roman letters indicate the best maxima found; italics denote other, lower maxima.

atom in C_2. We expect that such subtle inaccuracies in the practical search schemes will not significantly affect the results in MQSM studies.

MQSM and Tertiary-Structure Motifs

Robustness evaluation of global search techniques presents the inherent complexity of discerning whether found maxima are the certain, global maxima. Special cases of *a priori* known, global maxima are provided by the superposition of two identical molecular structures. Examples of those cases are included in Table IX and Table X. There, self-similarities were computed through the different global search schemes.

A more interesting case, where global maximizers could be inferred, occurs in the identification of tertiary-structure motifs, i.e., patterns of protein, secondary-structure elements in 3-D space.[19] One might expect that a protein segment, compared with the whole molecule, will maximize the similarity function when matching its original site in the chain.

The three reported structures of the amino-terminal domain of *N*-cadherin (NCD1)[20] provided our final test on the robustness of the present global optimizer. NCD1 complexed with Yb^{3+} produces a mixture of two closely related crystal types. One lattice, referenced as NCG01 in the Brookhaven Protein Data Bank (PDB), presents a $P6_322$ symmetry, while the other, referenced as NCH01, belongs to the isomorphic subgroup P321. The third structure, NCI01 in the PDB, is formed when complexing NCD1 with UO_2^{2+}. Its lattice type belongs to the $P2_12_12$ group. The Gly 40–Gly 49 segment in the NCD1 adopts an *unusual*, nearly helical structure that features a succession of β turns and a set of β-like hydrogen bonds. The differences in the spatial arrangement induced by the crystal packing are manifested in the values of the Carbó indices. These indices, which are defined as[1]

$$C_{AB} \equiv z_{AB}(z_{AA}z_{BB})^{-1/2}, \qquad (52)$$

provide a normalized quantification of similarity. The values of indices C_{AB} for the Gly 40–Gly 49 segments are reported in Table XII. In spite of the small magnitude crystal distortions, their impact in the indices is extreme, resulting in an electron cloud overlap of less than 25%. Changes in the spatial arrangement are observable in Figure 3, where the superposed fragments arising from the NCG01 and NCI01 lattices are displayed.

The segment Gly 40–Gly 49, extracted from lattice NCG01, was searched in the three protein structures following the first-level approach algo-

TABLE XII.
Carbó Indices for Three Crystal Structures of Segment Gly 40–Gly 49.

	NCG40 / 49	NCH40 / 49	NCI40 / 49
NCG40 / 49	1	0.252	0.186
NCH40 / 49		1	0.256
NCI40 / 49			1

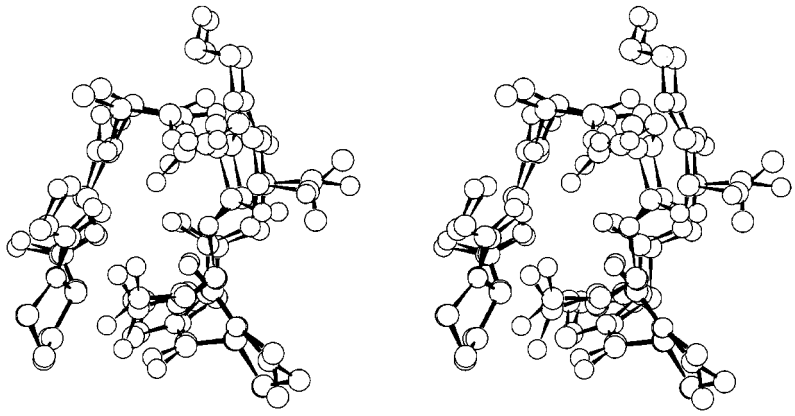


FIGURE 3. Stereoview of the superposed segments, Gly 40–Gly 49 in protein NCD1, from the crystal types $P6_322$ and $P2_12_12$. Pictured with Insight II (Biosym Technologies).

rithm. The *unusual*, quasi-β-helix motif associated with these segments infers that global maxima will overlap the two sequences Gly 40–Gly 49. However, because segments from different crystal lattices scarcely overlap, global maxima, in those cases, do not stand out against other local maxima in the function domain. Thus, such cases offer a valuable test of the robustness of the algorithm. Table XIII presents the values of the maxima found in the three cases, together with the ones coming from the comparison of the fragments. Observing them one realizes that global maxima were properly identified. The shift of approximately 19 au corresponds to the protein environment overlap. Global maximum achievement is easily manifested inspecting the translating–orienting pairs of atoms identifying the best estimate of a maximizer. In the three cases, the search provided pairs of equivalent atoms, that is, pairs of atoms having the same sequence number in the NCD1 protein. Note, finally, that the first-level algorithm is incomplete and approximate.

Computations were performed using a pure, promolecular Lorentzian-like approach. Thus, taking $z_{ab}(0)$ in function (38) as if a and b were free atoms avoids electron density computations. Table VII lists values of self-similarity at this approximate level for a possible accuracy comparison. Because overestimation in the Lorentzian-like similarity function is nearly constant along the function domain, one will expect accurate estimates for the relative magnitudes of C_{AB}. The search for the tertiary-structure motif required 5 h and 40 min on a single processor of an IBM 9076 SP2 supercomputer. The NCD1 protein has, depending on the lattice type and excluding hydrogen atoms, $\sim$800 atoms, while the Gly 40–Gly 49 segment has 62 atoms. The ASASim program was designed to deal with relatively small molecules and does not include speeding techniques, usually implemented in protein modeling packages. Then, one might expect an important reduction, possibly more than one order of magnitude, in the computation timings of very large molecules.

TABLE XIII.
Similarity Maxima for NCG01 Quasi-β-Helix Sequence Searched in Three Crystal Structures of Protein NCD1 and Maxima for Equivalent Segment–Segment Optimizations.

	z'_{AB}	z_{AB}	z'_{AB}	z_{AB}	
NCG	2901.227	2901.227	2882.378	2882.378	NCG40 / 49
NCH	703.585	744.417	684.940	725.835	NCH40 / 49
NCl	504.153	554.775	484.730	535.346	NCl40 / 49

Conclusions

Global optimizers are of paramount importance in the theory of molecular similarity[1] as well as in the QSAR practice.[2–4,6]

In this work we presented a limit solution to outline the path to a complete maximization of the MQSM, electron overlap, similarity function. Optimizing each estimate of a maximizer given in such a limit solution satisfies the theoretical interest of systematically identifying all maxima in the similarity function.

Moreover, the proposed *truncated* algorithms and the simplified similarity functions provide a rapid tool for practicable measures of similarity between molecules. These approximate search techniques have proved to be accurate enough and more robust than general, stochastic procedures. One also might expect encouraging results when using these algorithms as heuristic optimizers in molecular alignments based on other similarity functions.

Acknowledgments

This work was partly performed on the IBM SP2 array of the Centre de Supercomputació de Catalunya (CESCA) using a generous grant of computing time. One of us (P.C.) benefited from a fellowship (CIRIT OA/au BQF93/24) from the Catalan Government, while another of us (L.A.) benefited from a Ministerio de Educación y Ciencia fellowship.

References

1. (a) R. Carbó, L. Leyda, and M. Arnau, *Int. J. Quantum Chem.*, **17**, 1185 (1980); (b) R. Carbó and B. Calabuig, *Int. J. Quantum Chem.*, **42**, 1681 (1992); (c) R. Carbó and B. Calabuig, *Int. J. Quantum Chem.*, **42**, 1695 (1992); (d) R. Carbó, B. Calabuig, and E. Besalú, *Adv. Quantum Chem.*, **25**, 253 (1994); (e) R. Carbó, E. Besalú, J. Mestres, and M. Solà, *Topics Curr. Chem.*, **173**, 31 (1995); (f) R. Carbó and E. Besalú, In *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, R. Carbó, Ed., Kluwer Academic Publishers, Amsterdam, 1995, p. 3.
2. (a) A. Itai, N. Tomioka, M. Yamada, A. Inoue, and Y. Kato, In *3D QSAR in Drug Design. Theory, Methods and Applications*, H. Kubinyi, Ed., ESCOM Science Publishers B.V., Leiden, The Netherlands, 1993, p. 200; (b) C. G. Wermuth and T. Langer, in *3D QSAR in Drug Design. Theory, Methods and Applications*. H. Kubinyi, Ed., ESCOM Science Publishers B.V., Leiden, The Netherlands, 1993, p. 117; (c) A. K. Ghose, M. E. Logan, A. M. Treasurywala, H. Wang, R. Wahl, B. E. Tomczuk, M. R. Gowravaram, E. P. Jaeger, and J. J. Wendoloski, *J. Am. Chem. Soc.*, **117**, 4671 (1995); (d) D. M. Mottola, S. Laiter, V. J. Watts, A. Tropsha, S. D. Wyrick, D. E. Nichols, and R. B. Mailman, *J. Med. Chem.*, **39**, 289 (1996); (e) B. F. Thomas, I. B. Adams, S. W. Mascarella, B. R. Martin, and R. K. Razdan, *J. Med. Chem.*, **39**, 471 (1996).
3. S. C. Nyburg, *Acta Crystallogr.*, **B30**, 251 (1974).
4. (a) Y. C. Martin, M. G. Bures, and P. Willet, In *Reviews in Computational Chemistry*. K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, Inc., New York, 1990, p. 213; (b) Y. C. Martin, *J. Med. Chem.*, **35**, 2145 (1992); (c) P. Willet, In *Molecular Similarity in Drug Design*, P. M. Dean, Ed., Blackie Academic & Professional, London, 1995, p. 110; (d) J. S. Mason, In *Molecular Similarity in Drug Design*, P. M. Dean, Ed., Blackie Academic & Professional, London, 1995, p. 138.
5. (a) A. D. McLachlan, *Acta Crystallogr.*, **A28**, 656–657 (1972); (b) E. Gavuzzo, S. Pagliuca, V. Pavel, and C. Quagliata, *Acta Crystallogr.*, **B28**, 1968 (1972); (c) W. Kabsch, *Acta Crystallogr.*, **A32**, 922 (1976); (d) A. L. Mackay, *Acta Crystallogr.*, **A33**, 212 (1977); (e) A. D. McLachlan, *Acta Crystallogr.*, **A38**, 871 (1982).
6. (a) D. J. Danziger and P. M. Dean, *J. Theor. Biol.*, **116**, 215 (1985); (b) R. A. Lewis and P. M. Dean, *Proc. R. Soc. Lond.*, **B236**, 141 (1989); (c) M. T. Barakat and P. M. Dean, *J. Comput.-Aided Mol. Design*, **4**, 295 (1990); (d) M. T. Barakat and P. M. Dean, *J. Comput.-Aided Mol. Design*, **4**, 317 (1990); (e) M. T. Barakat and P. M. Dean, *J. Comput.-Aided Mol. Design*, **5**, 107 (1991); (f) M. C. Papadopoulos and P. M. Dean, *J. Comput.-Aided Mol. Design*, **5**, 119 (1991); (g) G. Diana, E. P. Jaeger, M. L. Peterson, and A. M. Treasurywala, *J. Comput.-Aided Mol. Design*, **7**, 325 (1993).
7. P. Constans and R. Carbó, *J. Chem. Inf. Comput. Sci.*, **35**, 1046 (1995).
8. P. Constans, L. Amat, X. Fradera, and R. Carbó, In *Advances in Molecular Similarity*, vol. 1, R. Carbó and P. G. Mezey, Eds., JAI Press Inc., Greenwich, CT, 1996, p. 187.
9. (a) K. Ruedenberg, R. C. Raffenetti, and D. Bardon, *Proceedings of the 1972 Boulder Conference on Theoretical Chemistry*, Wiley, New York, 1973, p. 164; (b) M. W. Schmidt and K. Ruedenberg, *J. Chem. Phys.*, **71**, 3951 (1979); (c) D. F. Feller and K. Ruedenberg, *Theor. Chim. Acta*, **52**, 231 (1979).
10. (a) K. Ruedenberg and W. H. E. Schwarz, *J. Chem. Phys.*, **92**, 4956 (1990); (b) P. Coppens, In *International Tables for Crystallography*, Kluwer Academic Publishers, Amsterdam, vol. B, 1992, p. 10; (c) P. Coppens and J. Becker, In *International Tables for Crystallography*. Kluwer Academic Publishers, Amsterdam, vol. C, 1992, p. 628.
11. F. S. Khul, G. M. Crippen, and D. K. Freisen, *J. Comput. Chem.*, **5**, 24 (1984).
12. (a) L. Piela, J. Kostrowicki, and H. A. Scheraga, *J. Phys. Chem.*, **93**, 3339 (1989); (b) J. Kostrowicki, L. Piela, B. J. Cherayil, and H. A. Sheraga, *J. Phys. Chem.*, **95**, 4113 (1991); (c) J. Pillardy, K. A. Olszewski, and L. Piela, *J. Phys. Chem.*, **96**, 4337 (1992); (d) J. Pillardy and L. Piela, *J. Phys. Chem.*, **99**, 11805 (1995).
13. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes. The Art of Scientific Computing*, Cambridge University Press, New York, 1992.
14. M. A. Wolfe, *Numerical Methods for Unconstrained Optimization*, Van Nostrand Reinhold Company Ltd., Berkshire, England, 1978, p. 89.

15. F. H. Allen and O. Kennard, *Chemical Design Automation News*, **8**, 31 (1993).
16. M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head–Gordon, C. Gonzalez, and J. A. Pople, *Gaussian 94, Revision B.3*, Gaussian, Inc., Pittsburgh, PA, 1995.
17. P. Constans and R. Carbó, *ASA Calculations v2.0*, Institute of Computational Chemistry, University of Girona, 1995.
18. J. A. Nelder and R. Mead, *Comput. J.*, **7**, 308 (1965).
19. P. Artymiuk, H. M. Grindley, A. B. Mackenzie, D. W. Rice, E. C. Ujah, and P. Willett, In *Molecular Similarity and Reactivity: From Quantum Chemical to Phenomenological Approaches*, R. Carbo, Ed., Kluwer Academic Publishers, Amsterdam, 1995, p. 123.
20. L. Shapiro, A. M. Fannon, P. D. Kwong, A. Thompson, M. S. Lehmann, G. Grübel, J. F. Legrand, J. Als–Nielsen, D. R. Colman, and W. A. Hendrickson, *Nature*, **374**, 327 (1995).